Advance Journal of Food Science and Technology 9(11): 827-831, 2015
DOI: 10.19026/ajfst.9.1638
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: March 28, 2015 Accepted: April 22, 2015 Published: September 25, 2015

Research Article

The Implementation of Speech Engine Based on Speex used in Foreign-food Interaction Learning System

Hui Hu
Henan College of Finance and Taxation, Henan, China

Abstract: With the continuous improvement and development of speech recognition technology, the numerous special purpose chips for food introduction speech recognition have been developed, thus, the practical products of speech recognition have been gradually appeared in the food market. This study will take the foundation of speech coding in Foreign-food Foreign-food speech system as the breakthrough point, by means of the interpretation of Speex, it discusses the design of the speech engine as well as the route of implementation so as to reduce tedious work in voice testing. Combined with the analysis of the basic structure of speech recognition in the Foreign-food system, it discusses the architecture of the speaker independent Foreign-food speech recognition learning system.

Keywords: Foreign-food, speex, speech coding, speech learning system

INTRODUCTION

Speech coding is the basic technology of digital speech transmission and storage, by means of the compressed digital; it can represent the speech signals and make the expression of these signals with the minimum number of the required bits. Compared with the stimulated voice, digital voice transmission and storage system of using speech coding technology, has the advantages of high reliability, strong anti-interference ability, easy to be quickly exchanged, easy for the realization of confidentiality, multiplexing, packaging as well as the advantage of low price, Savojim (1989). The compressed voice is used for transmission, which can reduce the required bandwidth of each route, thus it can transmit mote voice transmission in the same bandwidth; it can be used for storage, which can save space and improve the storage of speech length as well as reduce cost. With the development of computer technology, signal processing as well as the development of pattern recognition technology, speech recognition technology has been improved gradually, the fields of application have been more and more extensive, at the same time, a lot of speech recognition products have been appeared. Speech recognition products are widely used in voice dialing system, Foreign-food and Chinese translation system, intelligent toys controlling, intelligent home controlling system, smart phones, stock trading system, banking service system, medical intelligence service, the intelligent automobile navigation, industrial control and some other fields (Paulett and Langlotz, 2009).

MATERIALS AND METHODS

The interpretation of speex: Speex is a multi-mode, multi-rate, speech code, based on CELP algorithm, it can provide narrow band (Rabiner and Juang, 1993), wide band and ultra-wide band three speech codec modes, which are respectively corresponding to the speech signals with the bandwidth of 4 kHz (sampling rate is 8000), 8 kHz (sampling rate is 16000), 16 kHz (sampling rate is 32000). Among them, the narrow band speech coding only adopts narrow band sub-pattern coding; wide band speech can be divided into two sub bands, wide band speech adopts broadband sub-pattern coding, low band voice adopts narrow band sub-pattern coding; ultra-wide band will repeat decomposition two times which adopts the wide band sub-pattern coding twice and narrow band sub-pattern coding once (Pitton *et al*., 1994). As shown in Fig. 1. Thus, we can see, throughout the algorithm of Speex, it is composed by two types of sub-pattern encoding: narrow band sub-pattern and wide band sub-pattern.

The structure of the functional module of the foreign-food foreign-food interaction learning system: In this study, VOIP system adopts Speex coding. The whole system consists of two parts, the server side and client side. The server of the database saves the data of all registered users. Each user must firstly log on the server used by the client side and access to the list if online friends, then it can make a voice call to the online friends (Rosti and Gales, 2004). The call between the client sides adopts the mode of point to point. It can avoid the excessive delay of voice.

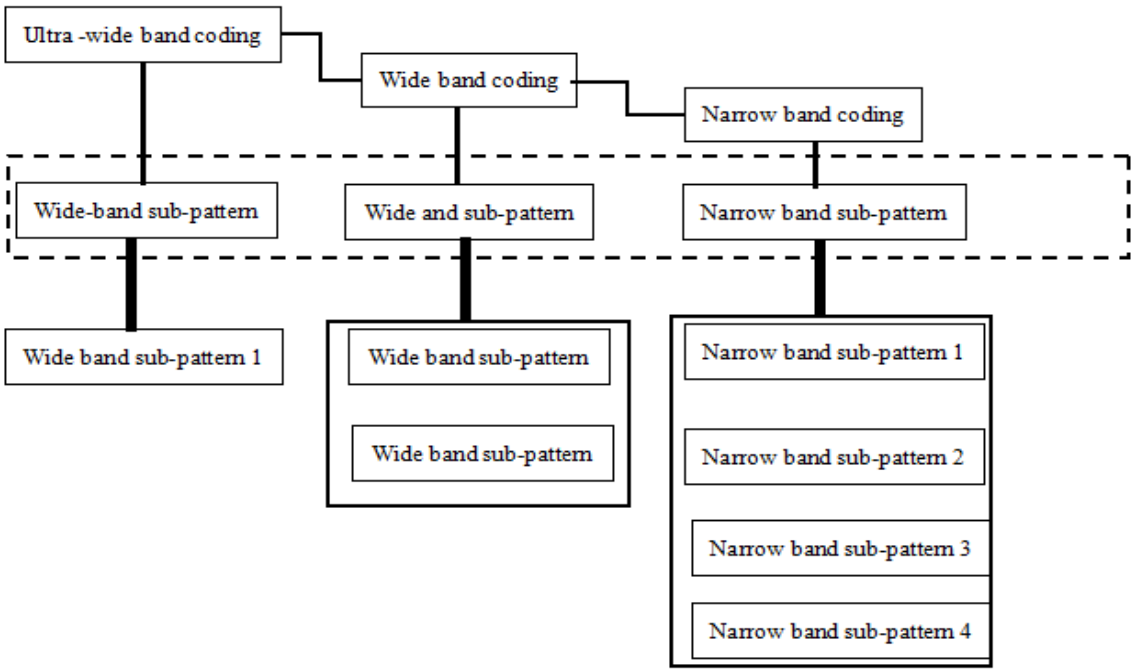

Fig. 1: Decomposition of coding mode

packet caused by the transmission of the server side Foreign-food Foreign-food interaction learning system refers to the application of using various advanced microprocessor with the realization of speech recognition technology in board-level or chip-level with software or hardware. Foreign-food Foreign-food interaction learning system is required to achieve the optimization of algorithm under the premise of ensuring the recognition effect as much as possible, so as to adapt to the characteristics of the Foreign-food platform with less storage resources and real-time (Tomio and Akira, 1997). The large vocabulary continuous Foreign-food Foreign-food interaction learning system with high performance in the advanced level of laboratory can represent today's advanced speech recognition technology. But because of the limitation of the Foreign-food platform in the aspects of resources and speed, the Foreign-food implementation has still not been mature. While, because the algorithm of small vocabulary speech commands recognition system is relatively simple, whose demand for resources is small and the rate of high recognition and robustness is rather high, which can meet the requirements of most applications, thus it has become the main focus of the Foreign-food application.

When the voice call begins, first of all, it should accept interface control information. This information can call the starting function of local voice and the starting function of the opposite terminal, judging whether it has opened coding thread or decoding thread, if it is not opened, then open it. The data after the coding thread is dealt with the data sending function, the data solved by decoding thread is dealt with the sound card to play.

When the voice call ends, first of all, it should accept interface control information. This information can call the termination function of local voice and the termination function of the opposite terminal, judging whether it has opened coding thread or decoding thread, if it is not opened, then open it.

Speech recognition is a pattern recognition in essence, but its voice signal is more complicated, plus its content is quite rich, so speech recognition is much complicated than the general pattern recognition, the Foreign-food Foreign-food interaction learning system mainly includes speech signal pre-treatment, endpoint detection, feature parameter extraction, pattern matching, reference template library and several other modules, the principle diagram of the system is shown in Fig. 2.

The basic structure of Foreign-food Foreign-food interaction learning system includes:

- Pre-treatment includes the collection of speech signal and the operation of the pre -emphasis, window adding and framing operation and so on.
- The endpoint detection can separate the speech signal effectively from the collected speech signals.
- Feature parameter extraction refers to extract key characteristic parameters from the signals that can reflect the characteristics of the speech signals.
- The training phase refers to the acquired feature parameter vector after the speech signal is input by the user, the pre-treatment of the speech signal, endpoint detection and feature extraction, which takes the characteristic parameters of each speech

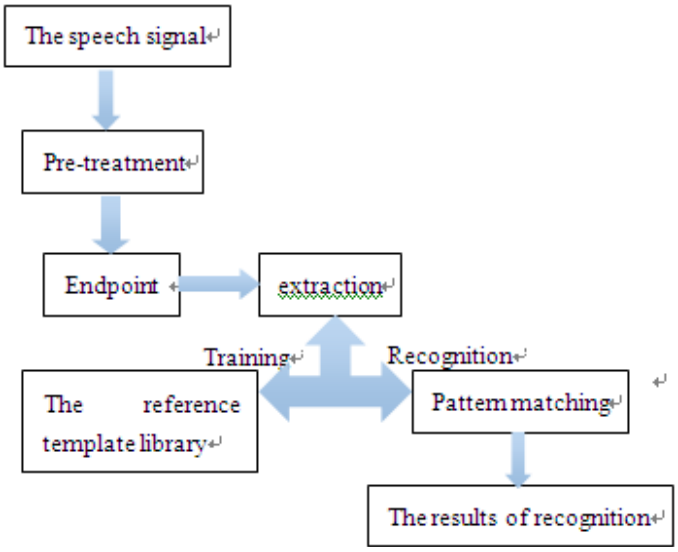

Fig. 2: Block diagram of Foreign-food Foreign-food interaction learning system

signal as the template and the formation of the reference template library.

- The recognition stage refers to contrasting the similarity between the feature parameter vector of the unknown speech signal and the reference template in the template library, taking the highest similarity model as the recognition results and then output.

Speech monitoring system of Foreign-food PI transmission can realize the localization of the processing of speech information, so as to improve the performance of the server. Each device can have access to the Internet with the service function, namely, for each speech equipment, it can be regarded as an independent network terminal, thereby it can greatly improve the quality and scope of monitoring.

Foreign-food IP speech equipment as well as CP are connected in the network, each speech equipment of the whole system and CP can be regarded as the network device that has a unique PI address, thus each equipment can be recognized by the address of network PI. For each speech equipment, the most basic function is to collect speech, playback, compress the code and decode, act as network interface, etc. Each speech equipment is equivalent to a speech acquisition and monitoring equipment, they are working under the remote monitoring of CP, which can complete the acquisition of data, compress the code and transmit the data. Remote CP can directly monitor the location of the speech device and store the speech information, when it is necessary, it can broadcast through the network in a single or multiple speech broadcasting equipment.

The design of the server side: The server side can not only save the user's information in addition to maintaining a database, but also can command and manage each client side. Once the server is started to start the service, it is started to listen for the requests of users. The server receives the message sent by the client side, firstly, it sends back the confirmed information; then it can establish a separate thread to deal with the received data. In this separate thread, according to the category of the received data it has the corresponding treatment.

The design of the client side: The function of the client side can be divided into two parts: one part is interacted with the server side, which can obtain the relevant information from the server; the other part can complete the communication between the different clients point to point. Among them, the second part is the core of VOIP system.

After starting the program of the client side, if there is local users' information, then it can load the local users' information and display the login window; if not, then it can display the user's registration window (in the login window, it also can choose the user's registration). Users can receive the request of voice call sent by friends, at the same time according to the user's operation, the client side can give response to friends, answer or reject.

RESULTS AND DISCUSSION

The hierarchical design of speech engine: The Inst Audio layer is located at the upper layer, involved with the interaction of interface; Voice Engine layer is in the middle layer, which is the middle part to play the function of calling different kinds of speech database; Speex base layer is codec library, which is the specific location that can implement the Speex codec as well as the process of mixing operation.

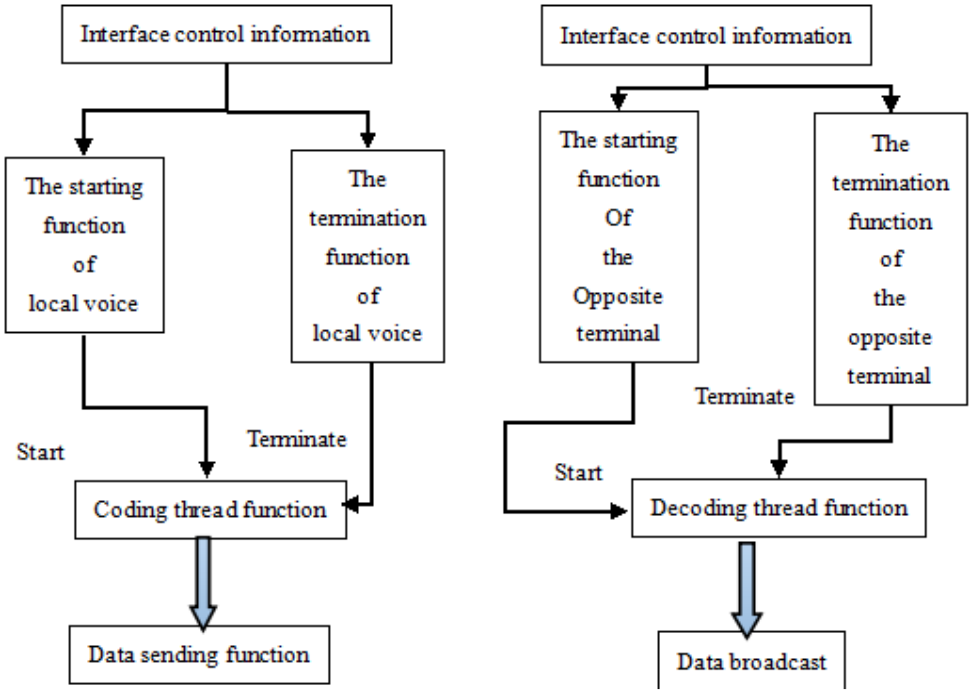

Fig. 3: The call graph of the relationship between components

The introduction of components:

The starting function of the local voice: As the entry function of the local voice in the conversation of each window, it has speech right, here, modifying the corresponding speech right of the local voice in chain P1 to be true, at the same time, finding out if there is the corresponding element of chain P2, if not, joining the corresponding element in chain P2, so as to check whether the thread is opened or not, if not, open the thread, otherwise skip it without processing. This function is in Voice Engine layer.

Encoding thread function: This function as the implementation of encoding function, namely encoding thread function, appeared as an independent running thread. This function is in Speex layer.

The termination function of the local voice: Such function is responsible for dealing with voice conversation between the local voice and either end of the opposite terminal when the local voice loses the speech rights. This function is in Voice Engine layer.

Data sending function: This function will be responsible for sending the compiled data by the callback method. This function is in Voice Engine layer.

The starting function of the opposite terminal: As the entry function in the conversation of each window, the opposite terminal has the speech right, her modifying the corresponding speech right of the opposite terminal in chain P1 to be true, at the same time, finding out if there is the corresponding element of chain P3, if not, joining the corresponding element in chain P3, so as to check whether the thread is opened or not, if not, open the thread, otherwise skip it without processing. This function is in Voice Engine layer.

Decoding thread function: This function as the implementation of decoding function, namely decoding thread function, appeared as an independent running thread. This function is in Speex layer.

Data receiving function: This function is responsible for dealing with the data sent by the opposite terminal, after acquiring the data, it can check whether the data has the corresponding data sub-chain in chain P4, if it has, add it at the end of the data sub-chain, if not, add the data sub-chain in chain P4. This function is in Voice Engine layer.

The relationship between components: The call graph of the relationship between components can be shown in Fig. 3.

CONCLUSION

In this study, the implementation of the speech engine based on Speex in Speaker independent Foreign-food interaction learning system is introduced in this study, which has many advantages compared with the speaker dependent isolated word Foreign-food interaction learning system, thus it can become the

main focus of researching the Foreign-food interaction learning system research as well as implementation. As for the Foreign-food platform, researching and developing the front-end processing module of special speech recognition can make it perform more complex speech front-end signal processing algorithms. And it has performed well at this stage under the environment. It has showed the efficiency of Speex as the voice codec, as well as the processing ability in voice communication. Foreign-food interaction learning system has a broad prospect of market application.

REFERENCES

Paulett, J.M., and C.P. Langlotz, 2009. Improving language models for radiology speech recognition. J. Biomed. Inform., 42: 53-58.

Pitton, J., L. Atlas and P. Loughlin, 1994. Aplications of possitive time-frequeney distributions to speeeh proeessing. IEEE T. Speech Audi. P., 2(4): 544-566.

Rabiner, L. and B. Juang, 1993. Fundamentals of Speech Recognition. Prentice-Hall, Englewood Cliff, New Jersey.

Rosti, A.V.I. and M.J.F. Gales, 2004. Factor analysed hidden Markov models for speech recognition. Comput. Speech Lang., 18: 181-200.

Savojim, H., 1989. A robust algorithm for accurate endpointing of speech. Speech Commun., 8: 45-60.

Tomio, T. and H. Akira, 1997. Speech recognition using the model structure determined by the genetic algorithm. Nonlinear Anal-Theor., 30: 2969-2979.